

ОСОБЕННОСТИ ПРИМЕНЕНИЯ МНОЖЕСТВЕННОГО КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Стребков Е.В., к.ф.-м.н., доцент,
Казанский федеральный университет, г. Казань
str9050258629@yandex.ru
Ахметзянов Д.И., студент,
Казанский федеральный университет, г. Казань
Сайтгареев Р.Д., студент,
Казанский федеральный университет, г. Казань

Аннотация. В этой статье рассматривается эффективность применения множественного коэффициента ранговой корреляции для комплексного корреляционного анализа при различных комбинациях факторов.

Ключевые слова: множественный коэффициент ранговой корреляции; комплексный корреляционный анализ.

FEATURES OF THE APPLICATION OF MULTIPLE CORRELATION COEFFICIENT

E.V. Strebkov, PhD, associate professor,
Kazan Federal University, Kazan
str9050258629@yandex.ru
D.I. Ahmetzyanov, student,
Kazan Federal University, Kazan
R.D. Saitgareev, student,
Kazan Federal University, Kazan

Abstract. This article discusses the efficacy of applying of plural coefficient of rank correlation for complex correlation analysis at different combinations of factors.

Keywords: multiple rank correlation coefficient; complex correlation analysis.

Для многих вузовских специальностей в курсе «Теория вероятностей и математическая статистика» рассматриваются корреляционные зависимости между несколькими признаками. Например, в [1] для трех признаков приведены достаточно сложные формулы вычисления частных и совокупного выборочных коэффициентов корреляции с использованием выборочных коэффициентов линейной корреляции по двум количественным признакам. При анализе корреляционных зависимостей между большим числом признаков студентам приходится использовать пакеты прикладных программ по математической статистике, что существенно ограничивает их применение при выполнении курсовых и выпускных квалификационных работ.

Более перспективным является применение множественного коэффициента ранговой корреляции [2]. Суть ранговых методов состоит в анализе отношений «больше-меньше» между реальными показателями изучаемых признаков.

Применение множественного коэффициента ранговой корреляции W проиллюстрируем на примере анализа для n объектов O_i зависимости между m признаками (факторами), значения которых приведены в Таблице 1, где x_{ij} – значение признака X_j на объекте O_i .

Таблица 1

Объекты	X_1	X_2	...	X_m
---------	-------	-------	-----	-------

O_1	x_{11}	x_{12}	\dots	x_{1m}
O_2	x_{21}	x_{22}	\dots	x_{2m}
\dots	\dots	\dots	\dots	\dots
O_n	x_{n1}	x_{n2}	\dots	x_{nm}

В Таблице 1 **раздельно для каждого признака X_j** ранжируются его значения x_{ij} , например, по убыванию. Соответствующие ранги r_{ij} значений x_{ij} приведены в Таблице 2.

Таблица 2

Объекты	X_1	X_2	\dots	X_m	R_i	R_i^2
O_1	r_{11}	r_{12}	\dots	r_{1m}	R_1	R_1^2
O_2	r_{21}	r_{22}	\dots	r_{2m}	R_2	R_2^2
\dots	\dots	\dots	\dots	\dots	\dots	\dots
O_n	r_{n1}	r_{n2}	\dots	r_{nm}	R_n	R_n^2

В Таблице 2 приведены суммы рангов соответствующих объектов O_i и их квадраты R_i^2

$$R_i = \sum_{j=1}^m r_{ij}.$$

Множественный коэффициент ранговой корреляции W имеет вид (случай несвязанных рангов)

$$W = \frac{12S}{m^2(n^3 - n)}, \quad (1)$$

где

$$S = \sum_{i=1}^n R_i^2 - \frac{1}{n} \left(\sum_{i=1}^n R_i \right)^2. \quad (2)$$

Коэффициент W принимает значения в пределах от 0 до 1, $W = 1$ означает наличие полной зависимости признаков X_i , при $W = 0$ – их независимость.

Значимость коэффициента W проверяется на основе критерия Пирсона χ^2 [1] с фактическим значением

$$\chi^2 = m(n + 1)W. \quad (3)$$

По заданному уровню значимости q и числу степеней свободы $k = n - 1$ из таблицы [1] выбирается критическое значение $\chi_{кр}^2$. При условии $\chi^2 > \chi_{кр}^2$ отклоняется гипотеза об отсутствии связи между признаками X_i , $i = 1, 2, \dots, m$.

Достоинством коэффициента W является также возможность без дополнительного ранжирования признаков X_i проанализировать наличие зависимостей для различных комбинаций от 2 до $m - 1$ этих признаков.

Существует связь коэффициента W с ранговым коэффициентом корреляции Спирмена для двух признаков [1]

$$\bar{\rho} = \frac{m * W - 1}{m - 1}, \quad (4)$$

где $\bar{\rho}$ - среднее для $m(m - 1)/2$ всевозможных пар коэффициентов корреляции Спирмена при исследовании зависимостей между m признаками.

Эффективность применения множественного коэффициента ранговой корреляции обусловлена его преимуществами:

- 1) применим для количественных и качественных признаков, которые поддаются ранжированию;
- 2) позволяет провести комплексный анализ наличия зависимостей при различных комбинациях рассматриваемых признаков;
- 3) является достаточно наглядным и не требует от пользователей дополнительных знаний по теории вероятностей и аналитической статистике;
- 4) применим для объектов, которые характеризуются одновременно количественными и качественными свойствами, что актуально при работе с большими данными.

Универсальность коэффициента W обеспечивает его применения для широкого круга задач из различных областей знаний, например, медико-биологических, социально-экономических, психолого-педагогических, информационных технологий.

Литература

1. Гмурман В. Е. Теория вероятностей и математическая статистика: Учебное пособие для вузов / В. Е. Гмурман. - М.: Высш. шк., 2003. – 479 с.
2. Холлендер М. Непараметрические методы статистики / М. Холлендер, Д. Вульф. – М.: Финансы и статистика, 1983. – 518с.